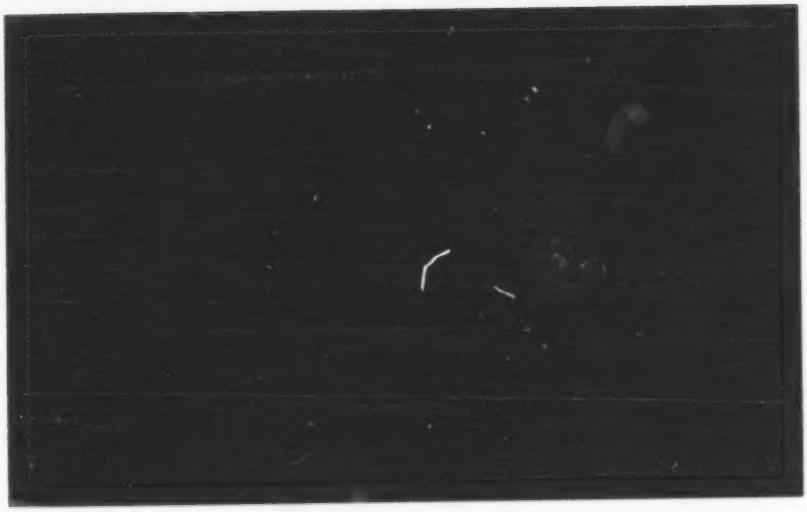
CENTRE DE RECHERCHE
SUR LES TRANSPORTS
(C.R.T.)
CENTRE FOR RESEARCH
ON TRANSPORTATION

UNIVERSITÉ DE MONTRÉAL
ÉCOLE DES HAUTES ÉTUDES COMMERCIALES
ÉCOLE POLYTECHNIQUE



CENTRE DE RECHERCHE SUR LES TRANSPORTS
UNIVERSITÉ DE MONTRÉAL
C.P. 6128, SUCCURSALE A.
MONTRÉAL
CANADA H3C 3J7

TÉLÉPHONE: (514) 343-7575
TÉLEX: (21) 05-561325
EMAIL: crt@crt.umontreal.ca
TÉLÉCOPIEUR (FAX): (514) 343-7121

CRT-733

Regression-Based Variance Estimators for the Error Components Model

by
Denis Bolduc*
and
Richard Laferrière**

* Département d'économique, Université Laval.
**Centre de recherche sur les transports, Université de Montréal.

We wish to thank Claude Montmarquette, Guy Lacroix and an anonymous referee for very helpful comments. Research support from the Social Sciences and Humanities Research Council of Canada and Quebec's FCAR grants are acknowledged.

Centre de recherche sur les transports —Publication #733
October 1990

91-00963/

Dépôt légal – 4e trimestre 1990

ABSTRACT

In this paper, we propose a class of variance estimators for the error components models which arises from applying the least squares principle. The suggested approach offers a new alternative when facing negative variance estimates and deals with unbalanced settings in a very natural way where no observation need be discarded. Finally, the least squares based variance estimators can handle general decompositions of the error term (such as nested structures and heteroscedastic components, for example) in a very straightforward way.

Keywords : error components model, variance estimators, two-way cross error model, two-way nested error model.

RÉSUMÉ

Dans ce texte, nous proposons un estimateur pour les variances du modèle à erreurs composées basé sur le principe des moindres carrés ordinaires. Cette approche offre de nouvelles perspectives aux problèmes d'estimations de variances négatives et s'applique même aux échantillons avec observations manquantes. De plus, nous définissons des estimateurs de variances pour différentes décompositions de l'erreur.

Mots-clés : modèle à erreurs composées, estimateurs de variances, modèle à erreurs croisées, modèle à erreurs imbriquées.

1. INTRODUCTION

The error components model offers many attractive ways to break down the error term in a regression. Unfortunately, difficulties related to the estimation of the variances sometimes limit its applicability. The existing variance estimators do not preclude the occurrence of negative variance estimates and they do not handle the problem of missing data (unbalanced data) properly. In this paper we propose a class of variance estimators which arises from applying the least squares principle. The suggested approach offers a new alternative when facing negative variance estimates and deals with unbalanced settings in a very natural way where no observation need be discarded. As noted in Baltagi and Li (1990), when dealing with incomplete panels (unbalanced data), the current practice is to drop extra observations in order to produce balanced data. As they indicate, this causes inefficiencies which make this practice unacceptable. Finally, as we shall see, the least squares based variance estimators can handle general decompositions of the error term (such as nested structures and heteroscedastic components, for example) in a very straightforward way.

In the next section, we first review the error components model. We then introduce two variance estimators and discuss their statistical properties. Section 3 describes the advantages of the new estimators and proposes some extensions. The last section contains the results of a Monte-Carlo experiment which enlarges the one by Baltagi (1981).

2. THE TWO-WAY ERROR COMPONENTS

The two-way error components model can be written as

$$y_{it} = \beta' x_{it} + v_{it}$$

$$\text{with, } v_{it} = \alpha_i + \lambda_t + u_{it}, \tag{1}$$

where, in a panel data setting, i, $(i = 1,\ldots,N)$, refers to an individual, a firm, a state or a cross-sectional unit and t, $(t = 1,\ldots,T)$, corresponds to a time period. With transportation data, i and t often represent a pair of zones in a region under study. Below, we consider that the data originate from a panel. It is assumed that β is a $(K \times 1)$ vector of fixed coefficients and x_{it} is a $(K \times 1)$ vector of FIXED regressors describing the mean of y_{it}. The components α_i is an individual time invariant cross-sectional random term assumed to be i.i.d. (independent and identically distributed) with a mean equal to zero and a variance σ_α^2. We denote this as: $\alpha_i \sim (0, \sigma_\alpha^2)$. Similarly, we assume that $\lambda_t \sim (0, \sigma_\lambda^2)$ and that $u_{it} \sim (0, \sigma_u^2)$. We further assume that these terms are uncorrelated with themselves and with each other. This is written as

$$E(\alpha_i \lambda_t) = E(\alpha_i u_{it}) = E(\lambda_t u_{it}) = 0 \quad \forall i,t.$$

Using a vector notation when considering all NT observations, the model becomes:

$$y = X\beta + \nu \tag{2}$$

with

$$\nu = (I_N \otimes e_T)\alpha + (e_N \otimes I_T)\lambda + u, \tag{3}$$

where α is $(N \times 1)$ and λ is $(T \times 1)$. e_l designates an l- dimensional unit vector and I_l an identity matrix of size l. The NT-dimensional vector y is formed by vertically stacking N vectors of size T, each one representing T observations of a given cross-sectional unit i, $i = 1,\ldots,N$. The X matrix and the ν vector are defined similarly. Under the assumptions previously made, $\nu \sim (0, \Phi)$, where:

$$\Phi = Var(\nu) = \sigma_\alpha^2(I_N \otimes e_T e_T') + \sigma_\lambda^2(e_N e_N' \otimes I_T) + \sigma_u^2(I_N \otimes I_T). \tag{4}$$

The first class of estimators suggested for the variances $\sigma_\alpha^2, \sigma_\lambda^2$, and σ_u^2 is based on the so-called "analysis of variance" estimators (AOVE). Assuming that the ν_{it} in equation (1) are observable, the AOVE estimators are computed as:

$$\ddot{\sigma}_\alpha^2 = \frac{1}{T(N-1)} \sum_{i=1}^{N} \left(\frac{1}{T}\nu_{i.} - \frac{1}{NT}\nu_{..}\right)^2 - \frac{1}{T}\ddot{\sigma}_u^2,$$

$$\ddot{\sigma}_\lambda^2 = \frac{1}{N(T-1)} \sum_{t=1}^{T} \left(\frac{1}{N}\nu_{.t} - \frac{1}{NT}\nu_{..}\right)^2 - \frac{1}{N}\ddot{\sigma}_u^2,$$

$$\ddot{\sigma}_u^2 = \frac{1}{(N-1)(T-1)} \sum_{i=1}^{N}\sum_{t=1}^{T} \left(\nu_{it} - \frac{1}{T}\nu_{i.} - \frac{1}{N}\nu_{.t} + \frac{1}{NT}\nu_{..}\right)^2. \tag{5}$$

The Wallace and Hussain (1969) (WALHUS) variance estimators are computed by replacing the unobservable ν with ordinary least squares (OLS) residuals. To provide a more efficient estimator for the variance components, Amemiya (1971) replaces the unobservable errors in AOVE by Least Squares with Dummy Variables (LSDV) residuals. A number of other estimators have been suggested since. They are well documented in econometrics textbooks such as Judge et al. (1985). In this paper, we suggest two new variance estimators which are based on OLS residuals. They are named LSVE1 and LSVE2. They both solve a regression problem which arises from rewriting the basic relationship $E(\nu\nu') = \Phi$ as follows:

$$\nu\nu' = \Phi + (\nu\nu' - \Phi), \tag{6}$$

where $(vv' - \Phi)$ designates the error term. In the final section of this paper, we describe the results of a Monte-Carlo experiment which compares the two proposed variance estimators with some of the major existing variance estimators. Following Baltagi (1981), the methods that we retain for comparison purposes in this study were developed by Wallace and Hussain (1969), Amemiya (1971), Nerlove (1971), Rao (1972) and Fuller and Battese (1974). To be consistent with the notation in Baltagi (1981), we call these methods WALHUS, AMEMIYA, NERLOVE, MINQUE and FUBA.

2.1 Results with the Unobservable Errors (Auxiliary Model)

Throughout the paper, vec(•) will designate the vertical concatenation of the columns of a matrix. Using this notation, one can write equation (6) as

$$\begin{aligned} r &= \text{vec}(I_N \otimes J_T)\sigma_\alpha^2 + \text{vec}(J_N \otimes I_T)\sigma_\lambda^2 + \text{vec}(I_N \otimes I_T)\sigma_u^2 + \varepsilon, \\ &= Z\theta + \varepsilon, \end{aligned} \tag{7}$$

where $r = \text{vec}(vv')$, $J_t = e_t e_t'$, $\theta = (\sigma_\alpha^2, \sigma_\lambda^2, \sigma_u^2)'$ and $Z = (\text{vec}(I_N \otimes J_T), \text{vec}(J_N \otimes I_T), \text{vec}(I_N \otimes I_T))$.

The N^2T^2-dimensional vector ε is the vectorized representation of the $(NT \times NT)$ error matrix $(vv' - \Phi)$. As one can easily verify, this error term has a vector mean equal to zero and a general variance-covariance matrix which incorporates heteroscedasticity as well as cross-correlations. This can be written as $E(\varepsilon) = 0$ and $Var(\varepsilon) = \Omega_{\varepsilon\varepsilon}$.

The GLS estimation is thus advisable on an efficiency standpoint. However, considering the size of the $\Omega_{\varepsilon\varepsilon}$ matrix, this only factor makes the implementation rather difficult. Furthermore, Taylor (1980) showed that a more efficient estimation of the variance components does not lead to an improved estimation of the parameters of interest. Therefore, we settle for simplicity and fall back on an OLS approach where θ^* denotes the solution to the normal equations

$$(Z'Z)\theta^* = Z'r.$$

Simple computations involving equations $(A10) - (A13)$ of the Appendix show that

$$\theta^* = \begin{pmatrix} N^2T & NT & NT \\ NT & NT^2 & NT \\ NT & NT & NT \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^{N} v_{i.}^2 \\ \sum_{t=1}^{T} v_{.t}^2 \\ \sum_{i=1}^{N}\sum_{t=1}^{T} v_{it}^2 \end{pmatrix}, \tag{8}$$

where the dot subscript indicates summation taken over it. For instance, $v_{i.}^2$ designates $\left(\sum_t v_{it}\right)^2$.

The closed form of the least squares estimators is given by:

$$\sigma_\alpha^{2*} = \frac{1}{NT(N-1)} \left(\sum_{i=1}^{N} v_{i.}^2 - \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it}^2 \right)$$

$$\sigma_\lambda^{2*} = \frac{1}{NT(T-1)} \left(\sum_{t=1}^{T} v_{.t}^2 - \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it}^2 \right) \qquad (9)$$

$$\sigma_u^{2*} = \frac{1}{NT(N-1)(T-1)} \left((NT-1) \sum_{i=1}^{N} \sum_{t=1}^{T} v_{it}^2 - (T-1) \sum_{i=1}^{N} v_{i.}^2 - (N-1) \sum_{t=1}^{T} v_{.t}^2 \right).$$

It is straightforward to show that the least squares estimator of θ is unbiased and consistent. Consistency follows from verifying that every diagonal element of $(Z'Z)^{-1}$ converges to 0 as N and T tend toward infinity (see, Amemiya, 1985, p. 95).

Although estimators in equation (9) seem at first to share some similarities with the estimators defined in equation (5), they in fact combine the components $v_{it}, v_{i.}, v_{.t}$ and $v_{..}$ in a very different way. One can actually verify that both sets of estimators always yield to different estimates. Now, it is not very interesting to compare analytically the proposed regression-based estimators with FUBA, MINQUE and NERLOVE since they all have a structure quite different from the functional forms given in equation (9).

2.2 The Two Least Squares Variance Estimators

We define $\hat{\sigma}_\alpha^2, \hat{\sigma}_\lambda^2$ and $\hat{\sigma}_u^2$ as the least squares variance estimators (LSVE1) where the OLS residuals $\hat{v}$ computed from equation (1) replace the unobservable v in equation (8).

Since X in equation (2) is fixed, it is then known that each element in $\hat{v}$ converges in distribution to v. Therefore LSVE1 is consistent and has the same asymptotic properties as the estimators in (8). In finite samples, however, the LSVE1 estimators are biased since

$$E(\hat{v}\hat{v}') = M\Phi M', \text{ where } M = (I_{NT} - X(X'X)^{-1}X'). \qquad (10)$$

To produce an unbiased estimation of θ, we suggest a second least squares variance estimator (LSVE2) where the right-hand side of equation (10) is the fixed part of the auxiliary regression, which then becomes:

$$\begin{aligned}\hat{r} &= \mathrm{vec}(M\Phi M) + w \\ &= \mathrm{vec}(M(I_N \otimes J_T)M)\sigma_\alpha^2 + \mathrm{vec}(M(J_N \otimes I_T)M)\sigma_\lambda^2 + \mathrm{vec}(M(I_N \otimes I_T)M)\sigma_u^2 + w, \\ &= Z(M)\theta + w,\end{aligned} \tag{11}$$

where $w = \mathrm{vec}(\hat{v}\hat{v}' - M\Phi M)$. Using the Appendix, the LSVE2 variance estimators can be proven to be equal to:

$$\begin{pmatrix} \tilde{\sigma}_\alpha^2 \\ \tilde{\sigma}_\lambda^2 \\ \tilde{\sigma}_u^2 \end{pmatrix} = (Z(M)'Z(M))^{-1} \begin{pmatrix} \sum_{i=1}^{N} & \hat{v}_{i.}^2 & \\ \sum_{t=1}^{T} & \hat{v}_{.t}^2 & \\ \sum_{i=1}^{N} & \sum_{t=1}^{T} & \hat{v}_{it}^2 \end{pmatrix}, \tag{12}$$

where

$$Z(M)'Z(M) = \begin{pmatrix} \sum_{i=1}^{N}\sum_{j=1}^{N}(e_T'M_{ij}e_T)^2 & \sum_{k=1}^{N} e_T'\left(\sum_{i=1}^{N} M_{ki}\right)\left(\sum_{j=1}^{N} M_{jk}\right)e_T & e_T'\left(\sum_{i=1}^{N} M_{ii}\right)e_T \\ \sum_{k=1}^{N} e_T'\left(\sum_{i=1}^{N} M_{ki}\right)\left(\sum_{j=1}^{N} M_{jk}\right)e_T & tr\left(\left(\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}\right)^2\right) & tr\left(\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}\right) \\ e_T'\left(\sum_{i=1}^{N} M_{ii}\right)e_T & tr\left(\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}\right) & NT-K \end{pmatrix},$$

where M_{ij} is a $(T \times T)$ submatrix of matrix M. Using the same arguments as for the LSVE1, one can verify that, in addition to being unbiased, this estimator denoted by $\tilde{\theta}$ is also consistent.

3. PROPERTIES OF LEAST SQUARES VARIANCE ESTIMATORS

The two OLS residual-based techniques used to estimate the variances offer several advantages over other existing methodologies. The major advantages are: 1) variance estimators can naturally be modified to handle missing observations (unbalanced data); 2) negative variance estimators can easily be dealt with; 3) several types of decompositions of the error term are possible; 4) variance estimators can easily be extended to the heteroscedastic error components case.

3.1 Missing Observations

LSVE1 and LSVE2 were derived assuming balanced data, (that is same number of T observations for each of the N cross-sectional units). However pooled cross-sections involve missing data very often in practice. As already mentioned, the current practice is to drop some observations in order

to produce a balanced setting. This leads to inefficiencies in the estimation of the parameters of interest. The most obvious way to tackle the problem of unbalanced data is to define a transformation matrix that maps an expanded balanced model into the current unbalanced setting. This technique is attractive because it uses all the available observations. However, it becomes untractable when the number of observations is large. The approach that we follow here also uses all active observations, but it is a lot more economical, computationally. We essentially augment the LSVE1 regression problem described in equation (7) with three dummy regressors to be defined below. Note that the approach that we suggest here only applies for the LSVE1 estimator. However, as the Monte-Carlo experiment will indicate later, the LSVE1 estimator is the most preferred one anyway. To derive the augmented LSVE1 regression equation, we first simplify the notation by rewriting equations (3) and (4) as

$$\nu = C_1\alpha + C_2\lambda + C_3 u, \tag{13}$$

and

$$\Phi = Var(\nu) = \sigma_\alpha^2 C_1 C_1' + \sigma_\lambda^2 C_2 C_2' + \sigma_u^2 I_{NT}. \tag{14}$$

Each C_i matrix can always be decomposed into a sum of two matrices $A_i + B_i$, where A_i corresponds to the non-missing observations of the C_i matrix and B_i corresponds to the missing observations of the C_i matrix. Using an obvious notation, this decomposition makes it possible to rewrite equation (14) as

$$\Phi = Var(\nu) = \sigma_\alpha^2 A_1 A_1' + \sigma_\lambda^2 A_2 A_2' + \sigma_\lambda^2 A_3 A_3' + \sigma_\alpha^2 F_1 + \sigma_\lambda^2 F_2 + \sigma_u^2 F_3. \tag{14'}$$

Define ν^* as the ν vector where each ν_{it} element corresponding to a missing observation is replaced by an arbitrary value that we denote by δ. With this in mind, we easily obtains

$$\Psi = Var(\nu^*) = \sigma_\alpha^2 A_1 A_1' + \sigma_\lambda^2 A_2 A_2' + \sigma_u^2 A_3 A_3'. \tag{15}$$

The Ψ matrix contains the error variances and covariances associated with the non-missing observations. These terms are assigned to the position that they share in the Φ matrix which corresponds to the complete $NT \times NT$ balanced setting. The least squares variance estimators that we propose here solve an OLS regression problem using a vectorization of Ψ as the fixed part. In this case, however, the LVSE estimators do not have a closed form. We now propose to rewrite this last equation into an equivalent form for which an explicit closed form solution exists. Using (14'), one can rewrite equation (15) as:

$$\Psi = \Phi - \sigma_\alpha^2 F_1 - \sigma_\lambda^2 F_2 - \sigma_u^2 F_3, \tag{16}$$

where the three F_i matrices share the same elements on their respective diagonals. To simplify the calculations of the variance estimators, we use the following alternative form of equation (16):

$$\Psi = \Phi - \sigma_\alpha^2(F_1 - F_3) - \sigma_\lambda^2(F_2 - F_3) - (\sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2)F_3. \tag{16'}$$

Applying the vec(•) operator to this last equation leads to the following final relationship:

$$\text{vec}(\Psi) = \sigma_\alpha^2 Z_1 + \sigma_\lambda^2 Z_2 + \sigma_u^2 Z_3 + \delta_1 D_1 + \delta_2 D_2 + \delta_3 D_3, \tag{17}$$

which contains the same Z_i regressors as in section 2, as well as three dummy vectors. It is now important to note that the three vectors $D_3 = F_3$, $D_1 = (\text{vec}(F_1) - D_3)$, and $D_2 = (\text{vec}(F_2) - D_3)$ are mutually orthogonal. The proposed LSVE1 estimator with missing data is defined by using equation (17) as the deterministic part of a least squares regression on $\hat{r}^* = \text{vec}(\hat{v}^* \hat{v}^{*\prime})$, where $\hat{v}_{it}^*$ which corresponds to a non-missing observation is the least squares residual $\hat{v}_{it}$ and where all $\hat{v}_{it}^*$ associated with a missing value is set to the constant value zero.

The D_i vectors have a very attractive property. On the condition that all $\hat{r}^*$ components associated with a missing observation are all set to the same constant value, their Boolean structure guarantees that the estimation in the amended balanced form is identical to the estimation obtained when using equation (15) directly. This last claim is based on the well-known result that in a regression, by using a Boolean variable that takes a non-zero value only for one observation is equivalent to dropping the observation from the sample. Here, the three Boolean D_i regressors play exactly this role with the data points corresponding to missing observations. We now provide the closed-form solution to the auxiliary regression problem which arises from using (17).

Assume that there are S missing observations and call d_1 and d_2 the number of ones in D_1 and D_2 respectively. It can be shown that:

$$\begin{aligned} d_1 &= NT(T-1) - \sum_{i=1}^{N} t_i(t_i - 1), \\ d_2 &= NT(N-1) - \sum_{t=1}^{T} n_t(n_t - 1), \end{aligned}$$

where t_i = number of periods associated with cross-sectional unit i,

n_t = number of sections associated with period t. (18)

Using this notation and the properties shared by the Boolean D_i dummy vectors, we can verify that the normal equations for this least squares problem are:

$$\begin{pmatrix} N^2T & NT & NT & d_1 & 0 & S \\ NT & NT^2 & NT & 0 & d_2 & S \\ NT & NT & NT & 0 & 0 & S \\ d_1 & 0 & 0 & d_1 & 0 & 0 \\ 0 & d_2 & 0 & 0 & d_2 & 0 \\ S & S & S & 0 & 0 & S \end{pmatrix} \begin{pmatrix} \hat{\sigma}_\alpha^2 \\ \hat{\sigma}_\lambda^2 \\ \hat{\sigma}_u^2 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{N} \hat{v}_{i.}^2 \\ \sum_{t=1}^{T} \hat{v}_{.t}^2 \\ \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \tag{19}$$

The closed form for the estimates can be computed as follows:

$$\hat{\sigma}_\alpha^2 = \frac{\sum_{i=1}^{N} \hat{v}_{i.}^2 - \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2}{(NT(N-1)-d_1)} \qquad \hat{\sigma}_\lambda^2 = \frac{\sum_{t=1}^{T} \hat{v}_{.t}^2 - \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2}{(NT(T-1)-d_2)}$$

$$\hat{\sigma}_u^2 = -\frac{\sum_{i=1}^{N} \hat{v}_{i.}^2}{(NT(N-1)-d_1)} - \frac{\sum_{t=1}^{T} \hat{v}_{.t}^2}{(NT(T-1)-d_2)} + k \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{v}_{it}^2, \tag{20}$$

where

$$k = \frac{N^3T^3 - T^3N(N+d_1+S) + TN(2S - NS - Nd_2) + S(d_2+d_1) + d_1d_2}{T^4N^3(N-1) - T^3N^2(N^2+N(S-1)+d_1-S) + T^2N(N^2(S-d_2)+N(d_1+d_2-S)+Sd_1) + TN(SNd_2 - S(d_1+d_2)+d_1d_2) - Sd_1d_2}.$$

3.2 Non-Negative Variances

The problem of negative variance estimates is well known in the error components literature. As argued in Baltagi (1981), negative estimates may occur because of model misspecification, (i.e. the generated model is a one-way error components model, but the estimated model is a two-way error components model), or because real variances are close to zero. Within our setting, we suggest two ways of coping with this problem.

One way is simply to impose non-negative constraints on the auxiliary regression that the variance estimators solve. This can easily be implemented through a constrained least squares estimation. It is however plausible that negative estimates of the variances are a consequence of using some poor error estimates. If that is the case, then a more interesting approach would be to run the auxiliary regression with observations corresponding to unreliable error estimates being flagged out. To compute the variance estimators, the idea is to apply the regression-based technique designed to handle unbalanced data. This methodology was proposed in the previous section. By retaining

only the subset of trustworthy error estimates, (unreliable observations could be replaced by their expected value 0 on the basis of some appropriate tests for outliers), one can improve the quality of the sample variance-covariance matrix being used.

3.3 Other Decompositions of the Error Term

The decomposition of the error term used in (1) is referred to as a two-way cross error model. One very interesting property of the estimation methods that we advocate for is that they can be applied to other types of error decompositions. For instance, Fuller and Battese (1973) and Searle (1971) have considered the estimation of the variances in a two-way nested error formulation.

With a balanced data set, the two-way nested error model can be written as

$$y_{its} = \beta' x_{its} + \alpha_i + \lambda_{it} + \nu_{its},$$

or using the matrix notation, as

$$y = X\beta + \nu$$

$$\text{with } \nu = (I_N \otimes J_{TS})\alpha + (I_{NT} \otimes J_S)\lambda + u$$

$$\text{where } i = 1,2,..,N; \quad t = 1,2,..,T; \quad s = 1,2,..,S. \tag{21}$$

In this model, α is $N \times 1$ and λ is $NT \times 1$. The nested model is different from the cross model since each category i has subcategory t instead of having category i crossed with category t. For example, Montmarquette and Mahseredjian (1988) used a two-way nested model to explain the scores y_{its} of a given student s in class t of a school i. The variance-covariance matrix of the random term is

$$\Phi = \sigma_\alpha^2 (I_N \otimes J_{TS}) + \sigma_\lambda^2 (I_{NT} \otimes J_S) + \sigma_u^2 I_{NTS}. \tag{22}$$

The LSVE1 variance estimators for the nested structure in (21) is easily shown to be

$$\begin{pmatrix}\hat{\sigma}_{\alpha}^2\\ \hat{\sigma}_{\lambda}^2\\ \hat{\sigma}_{u}^2\end{pmatrix}=\begin{pmatrix}NT^2S^2 & NTS^2 & NTS\\ NTS^2 & NTS^2 & NTS\\ NTS & NTS & NTS\end{pmatrix}^{-1}\begin{pmatrix}\sum_i \hat{v}_{i..}^2\\ \sum_i\sum_t \hat{v}_{it.}^2\\ \sum_i\sum_t\sum_s \hat{v}_{its}^2\end{pmatrix}, \tag{23}$$

with the following closed forms

$$\hat{\sigma}_{\alpha}^2 = \frac{\sum_i \hat{v}_{i..}^2 - \sum_i\sum_t \hat{v}_{it.}^2}{NTS^2(T-1)} \tag{24}$$

$$\hat{\sigma}_{\lambda}^2 = \frac{\left(\sum_i\sum_t \hat{v}_{it.}^2\right)(ST-1)}{(NS^3-NS^2)(T-1)T} - \frac{\sum_i \hat{v}_{i..}^2}{NTS^2(T-1)} - \frac{\sum_i\sum_t\sum_s \hat{v}_{its}^2}{NTS(S-1)} \tag{25}$$

$$\hat{\sigma}_{u}^2 = \frac{\sum_i\sum_t\sum_s \hat{v}_{its}^2}{NT(S-1)} - \frac{\sum_i\sum_t \hat{v}_{it.}^2}{NTS(S-1)}. \tag{26}$$

3.4 Estimation of Variances in Heteroscedastic Error Components Models

The heteroscedastic two-way error components model has been suggested by Mazodier and Trogon (1978). Some new results, irrelevant for the variance estimation, have also been obtained in Baltagi and Griffin (1988) for a one-way version of the model. In this section, we extend the LSVE estimators to the heteroscedastic case.

Recalling equation (1), the two-way heteroscedastic error components model is formulated as

$$y_{it} = \mu + \beta' x_{it} + v_{it}$$

$$v_{it} = \alpha_i + \lambda_t + u_{it}$$

where the hypotheses and the notation are unchanged except for the distributional assumptions which become $\alpha_i \sim (0, \sigma_{\alpha i}^2)$ $\lambda_t \sim (0, \sigma_{\lambda t}^2)$ and $u_{it} \sim (0, \sigma_u^2)$. In matrix notation, equations (2) and (3) still hold where now, $\alpha \sim (0, \Sigma_\alpha)$ with $\Sigma_\alpha = \text{diag}(\sigma_{\alpha 1}^2, \sigma_{\alpha 2}^2, \ldots, \sigma_{\alpha N}^2)$ and $\lambda \sim (0, \Sigma_\lambda)$ with $\Sigma_\lambda = \text{diag}(\sigma_{\lambda 1}^2, \sigma_{\lambda 2}^2, \ldots, \sigma_{\lambda T}^2)$.

Using these results, $\Phi = var(v)$ can now be written as

$$\begin{aligned}\Phi &= (I_N \otimes e_T)\Sigma_\alpha(I_N \otimes e_T') + (e_N \otimes I_T)\Sigma_\lambda(e_N' \otimes I_T) + \sigma_u^2(I_N \otimes I_T)\\ &= (\Sigma_\alpha \otimes J_T) + (J_N \otimes \Sigma_\lambda) + \sigma_u^2(I_N \otimes I_T)\\ &= \sum_{i=1}^{N} \sigma_{\alpha i}^2(E_{ii}^N \otimes J_T) + \sum_{t=1}^{T} \sigma_{\lambda t}^2(J_N \otimes E_{tt}^T) + \sigma_u^2(I_N \otimes I_T),\end{aligned} \tag{27}$$

where E_{ll}^{L} is an elementary matrix of order L with zeros everywhere and a one in position ll. The last expression in (27) makes it easy to define the corresponding LSVE1 and LSVE2 estimators. These estimators are based on Z and $Z(M)$ matrices with $(N+T+1)$ regressors. When referring to previous sections and the Appendix, the evaluation of those matrices is manageable.

4. A MONTE CARLO STUDY

In this section, we propose to enlarge the Monte Carlo study considered in Baltagi (1981) by adding the two estimators proposed in Section 2 for the standard two-way error components model. We wish to analyze the finite properties and the relative performance of estimated generalized least squares estimators of the coefficients of the fixed parts of the model and the performance of the estimators of the variances.

The model used by Baltagi is the following:

$$y_{it} = \mu + \beta x_{it} + \alpha_i + \lambda_t + u_{it}. \tag{28}$$

The subsequent values assumed for the parameters are:

$\mu = 5$, $\beta = .5$, $N = 25$, $T = 10$,

$\sigma_\alpha^2 + \sigma_\lambda^2 + \sigma_u^2 = 20$,

$B_{xx} = 832.196$, $C_{xx} = 494.883$, $W_{xx} = 100.198$, where

$B_{xx} = T \sum_i (\bar{x}_{i.} - \bar{x}_{..})^2$, $C_{xx} = N \sum_t (\bar{x}_{.t} - \bar{x}_{..})^2$,

$W_{xx} = \sum_i \sum_t (x_{it} - \bar{x}_{i.} - \bar{x}_{.t} + \bar{x}_{..})^2$,

where, as usual, B_{xx}, W_{xx} and C_{xx} denote the between, within and total sum of squares, respectively. The computations are made for several combinations of $\rho = \sigma_\alpha^2/\sigma^2$ and $\omega = \sigma_\lambda^2/\sigma^2$. The other estimators of the variances which appear in Table 1 are those of: Amemya (1971), Fuller and Battese (1974), Nerlove (1971), Rao (1972) and, Wallace and Hussain (1969).

The underlined values in Table 1 are associated with the estimation technique producing the lowest MSE for at least two variances or with the lowest total MSE for the three variances. The estimator of Fuller and Battese is preferred 16 times, followed by LSVE1 (5 times); the estimator of Wallace and Hussain is favored in 4 cases and the estimator of Nerlove only once.

One result of Baltagi's paper is that a smaller MSE estimator for the variances has no consequences on the MSE estimator for the parameters of the fixed part of the model. This result was proven analytically by Taylor (1980) in a one-way error components model. By comparing the results for the relative MSE of the generalized least squares estimators for the coefficients in Table 2 with the results in Table 1, we obtain a similar conclusion. From Table 2, the LSVE1 estimator is preferred in 12 cases, the estimator of Fuller and Battese is favored 6 times, Amemya's estimator 6 times and the techniques LSVE2 and WALHUS 1 time each. These results show that although the LSVE1 estimator is not the best in terms of the estimation of the variances, it largely dominates all the other techniques for the EGLS estimation of the β coefficients. From Table 2, one can also note that LSVE1 is preferred for very different combinations of $\sigma^2_\mu, \sigma^2_\lambda$, and σ^2_u values, which shows that this estimator performs better for a vast range of values of parameters.

TABLE 1 : MEAN SQUARE ERRORS OF THE ESTIMATORS

		AMEMIYA			LSVE1			LSVE2			FUBA			NERLOVE			RAO			WALHUS		
ρ	ϖ	σ_α^2	σ_1^2	σ_ν^2	σ_α^2	σ_1^2	σ_ν^2	σ_α^2	σ_1^2	σ_ν^2	σ_α^2	σ_1^2	σ_ν^2	σ_α^2	σ_1^2	σ_ν^2	σ_α^2	σ_1^2	σ_ν^2	σ_α^2	σ_1^2	σ_ν^2
0.00	0.00	1.509	0.638	3.644	0.348	0.130	3.583	0.392	0.154	3.618	0.378	0.150	3.600	8.237	1.867	10.803	1.505	0.643	3.663	0.364	0.137	3.614
	0.01	1.353	0.826	3.805	0.339	0.212	3.708	0.364	0.261	3.741	0.351	0.236	3.756	7.682	1.784	10.734	1.352	0.826	3.805	0.342	0.230	3.737
	0.20	1.250	6.531	2.384	0.471	4.319	2.583	0.263	5.178	2.352	0.232	4.752	2.372	5.317	5.027	6.744	1.245	6.518	2.409	0.253	4.625	2.349
	0.40	0.999	19.323	1.283	1.058	14.236	2.106	0.222	17.190	1.264	0.132	15.624	1.252	2.932	13.958	3.666	0.998	19.375	1.284	0.226	15.341	1.263
	0.60	0.903	36.274	0.630	1.935	28.982	2.412	0.200	35.433	0.594	0.057	32.760	0.594	1.156	28.047	1.632	0.902	36.304	0.627	0.223	31.863	0.595
	0.80	0.777	62.257	0.183	3.346	47.831	3.477	0.306	59.165	0.164	0.014	53.746	0.157	0.309	48.246	0.409	0.773	62.271	0.184	0.399	52.348	0.167
0.01	0.00	1.507	0.592	3.677	0.414	0.130	3.560	0.441	0.155	3.597	0.429	0.148	3.626	7.565	1.605	10.361	1.509	0.596	3.665	0.418	0.138	3.592
	0.01	1.392	0.735	3.433	0.407	0.202	3.383	0.431	0.241	3.397	0.410	0.222	3.404	7.465	1.643	9.986	1.390	0.742	3.440	0.407	0.213	3.393
	0.20	1.208	5.965	2.577	0.591	4.181	2.774	0.326	4.927	2.531	0.294	4.514	2.536	4.801	4.878	6.671	1.195	6.000	2.583	0.306	4.405	2.529
	0.40	1.163	16.037	1.296	1.068	13.160	2.055	0.250	15.230	1.274	0.178	14.174	1.268	2.868	12.916	3.713	1.160	16.047	1.285	0.248	13.870	1.273
	0.60	0.966	37.429	0.598	2.015	29.569	2.412	0.254	36.747	0.579	0.087	34.273	0.575	1.210	29.326	1.597	0.961	37.512	0.600	0.273	33.009	0.580
	0.80	0.842	70.914	0.157	3.378	55.970	3.445	0.337	67.530	0.145	0.027	61.627	0.137	0.300	55.603	0.385	0.838	70.943	0.157	0.444	60.413	0.148
0.20	0.00	4.020	0.567	2.246	2.437	0.138	2.231	2.598	0.113	2.189	2.433	0.098	2.188	6.801	1.147	6.509	4.035	0.571	2.247	2.445	0.100	2.186
	0.01	4.063	0.661	2.388	2.571	0.213	2.362	2.651	0.184	2.342	2.509	0.162	2.352	6.720	1.090	6.753	4.032	0.667	2.390	2.518	0.161	2.339
	0.20	4.286	5.576	1.308	2.771	4.144	1.594	2.563	4.419	1.261	2.387	4.068	1.273	5.051	4.024	3.774	4.281	5.568	1.307	2.391	3.989	1.259
	0.40	3.276	15.864	0.611	3.097	12.077	1.617	2.030	14.008	0.602	1.816	12.855	0.594	2.794	11.554	1.728	3.270	15.846	0.616	1.900	12.511	0.600
	0.60	2.938	39.058	0.183	4.248	30.773	2.343	1.841	37.198	0.159	1.607	33.853	0.157	1.648	29.565	0.432	2.931	39.115	0.184	1.712	33.098	0.160
0.40	0.00	9.558	0.459	1.389	7.327	0.208	1.452	7.816	0.084	1.349	7.413	0.055	1.355	9.265	0.625	3.898	9.540	0.463	1.395	7.386	0.077	1.347
	0.01	9.242	0.562	1.263	7.134	0.263	1.360	7.507	0.139	1.237	7.097	0.103	1.223	8.588	0.665	3.664	9.235	0.564	1.263	7.115	0.124	1.235
	0.20	8.324	5.837	0.618	6.421	4.175	1.213	6.369	4.268	0.588	5.997	3.877	0.580	6.715	3.808	1.622	8.332	5.853	0.618	5.950	3.782	0.589
	0.40	7.978	17.108	0.178	7.347	13.668	1.559	6.150	15.483	0.158	5.683	14.087	0.153	5.745	12.360	0.416	7.971	17.164	0.178	5.703	13.791	0.160
0.60	0.00	16.417	0.421	0.581	12.594	0.323	0.828	13.806	0.067	0.538	13.013	0.025	0.536	13.304	0.279	1.609	16.388	0.424	0.588	13.013	0.069	0.539
	0.01	14.788	0.500	0.630	12.293	0.371	0.865	13.206	0.100	0.595	12.715	0.061	0.597	12.481	0.292	1.643	14.789	0.505	0.634	12.654	0.094	0.595
	0.20	14.691	5.366	0.173	12.675	4.302	0.967	12.574	4.357	0.157	11.937	3.828	0.152	11.499	3.433	0.424	14.668	5.374	0.173	11.899	3.833	0.157
0.80	0.00	25.227	0.332	0.183	20.292	0.489	0.619	22.549	0.049	0.169	21.202	0.006	0.163	20.088	0.065	0.437	25.221	0.336	0.184	21.175	0.057	0.169
	0.01	23.875	0.384	0.153	20.022	0.537	0.594	21.662	0.076	0.137	20.817	0.027	0.133	19.783	0.082	0.384	23.867	0.385	0.152	20.735	0.079	0.137

TABLE 2 : MSE OF 2 STEPS EGLS / MSE OF GLS

		AMEMIYA		LSVE1		LSVE2		FUBA		NERLOVE		RAO		WALHUS	
ρ	ω	μ	β	μ	β	μ	β	μ	β	μ	β	μ	β	μ	β
0.00	0.00	1.027	1.027	1.017	1.018	1.018	1.020	1.015	1.017	1.398	1.417	1.027	1.027	1.017	1.018
	0.01	1.035	1.040	1.024	1.027	1.023	1.029	1.023	1.028	1.256	1.283	1.036	1.041	1.022	1.027
	0.20	1.065	1.083	1.038	1.047	1.046	1.057	1.063	1.077	1.367	1.462	1.064	1.082	1.054	1.067
	0.40	1.009	1.048	1.006	1.022	1.004	1.023	1.006	1.020	1.195	1.391	1.009	1.048	1.009	1.032
	0.60	1.057	1.057	1.003	1.008	1.011	1.023	1.009	1.021	1.179	1.297	1.057	1.058	1.017	1.033
	0.80	1.060	1.226	1.000	1.001	1.008	1.060	0.998	1.002	1.072	1.279	1.060	1.224	1.015	1.103
0.01	0.00	1.021	1.024	1.021	1.022	1.024	1.026	1.018	1.020	1.277	1.291	1.021	1.024	1.022	1.024
	0.01	1.022	1.027	1.005	1.006	1.011	1.012	1.010	1.011	1.266	1.281	1.022	1.026	1.008	1.009
	0.20	1.047	1.048	1.028	1.024	1.030	1.027	1.034	1.032	1.226	1.300	1.046	1.047	1.033	1.031
	0.40	1.044	1.070	0.999	1.012	1.018	1.040	1.019	1.040	1.281	1.433	1.044	1.069	1.023	1.047
	0.60	1.024	1.074	1.006	1.005	1.018	1.044	1.004	1.017	1.135	1.353	1.024	1.073	1.022	1.058
	0.80	1.053	1.233	1.006	1.017	1.009	1.060	1.001	1.015	1.076	1.285	1.052	1.232	1.019	1.098
0.20	0.00	1.089	1.096	1.042	1.048	1.070	1.084	1.039	1.049	1.290	1.327	1.089	1.097	1.071	1.085
	0.01	1.081	1.102	1.034	1.040	1.050	1.064	1.036	1.047	1.215	1.252	1.080	1.101	1.047	1.060
	0.20	1.009	1.010	1.076	1.095	1.047	1.058	1.055	1.067	1.111	1.132	1.010	1.010	1.047	1.058
	0.40	0.992	0.998	1.072	1.093	1.040	1.054	1.041	1.059	1.055	1.079	0.992	0.998	1.040	1.054
	0.60	0.987	0.995	1.089	1.154	1.002	1.019	1.002	1.021	1.003	1.030	0.986	0.996	1.000	1.016
0.40	0.00	1.182	1.197	1.036	1.042	1.086	1.104	1.061	1.072	1.416	1.473	1.184	1.199	1.097	1.116
	0.01	1.107	1.135	1.040	1.051	1.058	1.084	1.046	1.062	1.252	1.302	1.109	1.137	1.062	1.088
	0.20	1.025	1.030	1.100	1.113	1.056	1.059	1.057	1.064	1.088	1.095	1.026	1.030	1.053	1.057
	0.40	1.018	1.015	1.098	1.120	1.041	1.049	1.047	1.058	1.035	1.043	1.019	1.015	1.042	1.051
0.60	0.00	1.209	1.313	0.999	1.007	1.092	1.147	1.036	1.056	1.361	1.552	1.210	1.313	1.116	1.181
	0.01	1.130	1.159	1.024	1.035	1.091	1.118	1.044	1.054	1.236	1.287	1.131	1.160	1.099	1.126
	0.20	0.999	1.002	1.071	1.111	1.028	1.043	1.029	1.046	1.022	1.032	0.999	1.002	1.028	1.044
0.80	0.00	1.340	1.602	1.003	1.001	1.048	1.079	1.009	1.021	1.239	1.435	1.342	1.606	1.084	1.155
	0.01	1.141	1.205	1.080	1.129	1.045	1.088	1.010	1.032	1.095	1.167	1.140	1.204	1.043	1.081

5. CONCLUSION

In this paper, we used the least squares principle to estimate the variances of the error components model. This principle is useful since it allows the treatment of non-negative estimators, generalized random components, unbalanced data and heteroscedastic components. Moreover, this principle may be applied to a wide variety of decompositions of the error term. In addition to these theoretical advantages, our Monte-Carlo study shows that this principle, applied to a standard two-way cross error components model with fixed regressors, leads to a more efficient generalized least squares estimation of the coefficients of the fixed part.

6. BIBLIOGRAPHY

Amemiya, T. (1971), *The Estimation of Variances in a Variance-Components Model*, International Economic Review, 12, 1-13.

Baltagi, B.H. (1981), *Pooling: An Experimental Study of Alternative Testing and Estimation Procedures in a Two-Way Error Component Model*, Journal of Econometrics 17, 21-49.

Baltagi, B.H. and Q. Li (1990), *A Comparison of Variance Components Estimators Using Balanced versus Unbalanced Data*, Econometric Theory, vol.6, no. 2, 283-285.

Baltagi, B.H. and J.M. Griffin (1988), *A Generalized Error Component Model with Heteroscedastic Disturbances*, International Economic Review, 29, 745-752.

Fuller, W.A. and G.E. Battese (1973), *Transformations for Estimation of Linear Models with Nested Error Structure*, Journal of the American Statistical Association, vol.68, number 343, 626-632.

Fuller, W.A. and G.E. Battese (1974), *Estimation of Linear Models with Crossed-Error Structure*, Journal of Econometrics 2, 67-78.

Judge et al. (1985), *The Theory and Practice of Econometrics*, John Wiley and Sons, 2nd edition.

Mazodier, P. and A. Trognon (1978), *Heteroscedasticity and Stratification in Error Components Models*, Annales de l'Insee, 30/31, 451-482.

Montmarquette, C., and S. Mahseredjian (1987), *Does School Matter: An Application of a Two-Way Nested Error Component Model*, forthcoming in Journal of Applied Econometrics.

Nerlove, M. (1971a), *Further Evidence on the Estimation of Dynamic Economic Relations from a Time Series of Cross Sections*, Econometrica, 39, 359-382.

Nerlove, M. (1971b), *A note on Error Components Models*, Econometrica, 39, 383-396.

Rao, C.R. (1972), *Estimation of Heteroscedastic Variances in Linear Models*, Journal of the American Statistical Association, 67, 112-115.

Searle, S.R. (1971), *Topics in Variance Components Estimation*, Biometrics, 27, 1-76.

Taylor, W.E. (1980), *Small Sample Considerations in Estimation From Panel Data*, Journal of Econometrics, 13, 203-223.

Wallace, T.D. and A. Hussain (1969), *The Use of Error Components Models in Combining Cross Section with Time Series Data*, Econometrica, 37, 55-72.

Wansbeek, T.J. and A. Kapteyn (1982), *A Class of Decompositions of the Variance-Covariance Matrix of a Generalized Error Components Model*, Econometrica, 50, 713-724.

7. APPENDIX Derivation of the LSVE2 estimator

PRELIMINARY RESULTS

In the paper, the $(NT \times 1)$ error vector ν, displayed in equation (2) can be taken to originate from the relationship $\nu = \text{vec}(N')$, where

$$N = \begin{bmatrix} \nu_{11} & \nu_{12} & \cdots & \nu_{1T} \\ \nu_{21} & \nu_{22} & \cdots & \nu_{2T} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \nu_{N1} & \nu_{N2} & \cdots & \nu_{NT} \end{bmatrix}. \qquad (A1)$$

Partitioning the M matrix into N^2 $(T \times T)$ matrices, one obtains

$$M(I_N \otimes J_T) = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1N} \\ M_{21} & M_{22} & \cdots & M_{2N} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ M_{N1} & M_{N2} & \cdots & M_{NN} \end{bmatrix} \begin{bmatrix} J_T & 0 & \cdots & 0 \\ 0 & J_T & \cdots & 0 \\ . & . & \cdots & . \\ . & . & \cdots & . \\ 0 & 0 & \cdots & J_T \end{bmatrix}$$

$$= \begin{bmatrix} M_{11}J_T & M_{12}J_T & \cdots & M_{1N}J_T \\ M_{21}J_T & M_{22}J_T & \cdots & M_{2N}J_T \\ . & . & \cdots & . \\ . & . & \cdots & . \\ M_{N1}J_T & M_{N2}J_T & \cdots & M_{NN}J_T \end{bmatrix} \qquad (A2)$$

and

$$M(J_N \otimes I_T) = \begin{bmatrix} M_{11} & M_{12} & \cdots & M_{1N} \\ M_{21} & M_{22} & \cdots & M_{2N} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ M_{N1} & M_{N2} & \cdots & M_{NN} \end{bmatrix} \begin{bmatrix} J_{11}I_T & J_{12}I_T & \cdots & J_{1N}I_T \\ J_{21}I_T & J_{22}I_T & \cdots & J_{2N}I_T \\ . & . & \cdots & . \\ . & . & \cdots & . \\ J_{N1}I_T & J_{N2}I_T & \cdots & J_{NN}I_T \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} J_{i1}M_{1i} & \sum_{i=1}^{N} J_{i2}M_{1i} & \cdots & \sum_{i=1}^{N} J_{iN}M_{1i} \\ \sum_{i=1}^{N} J_{i1}M_{2i} & \sum_{i=1}^{N} J_{i2}M_{2i} & \cdots & \sum_{i=1}^{N} J_{iN}M_{2i} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \sum_{i=1}^{N} J_{i1}M_{Ni} & \sum_{i=1}^{N} J_{i2}M_{Ni} & \cdots & \sum_{i=1}^{N} J_{iN}M_{Ni} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{N} M_{1i} \\ \sum_{i=1}^{N} M_{2i} \\ . \\ . \\ \sum_{i=1}^{N} M_{Ni} \end{bmatrix} [I_T \quad I_T \quad \cdots \quad I_T]. \qquad (A3)$$

- MATRIX $Z(M)'Z(M)$:

The elements of this matrix can be computed as follows:

$tr(M(I_N \otimes J_T)M(I_N \otimes J_T))$

$$= tr\left(\begin{bmatrix} M_{11}J_T & M_{12}J_T & \cdots & M_{1N}J_T \\ M_{21}J_T & M_{22}J_T & \cdots & M_{2N}J_T \\ . & . & \cdots & . \\ . & . & \cdots & . \\ M_{N1}J_T & M_{N2}J_T & \cdots & M_{NN}J_T \end{bmatrix} \begin{bmatrix} M_{11}J_T & M_{12}J_T & \cdots & M_{1N}J_T \\ M_{21}J_T & M_{22}J_T & \cdots & M_{2N}J_T \\ . & . & \cdots & . \\ . & . & \cdots & . \\ M_{N1}J_T & M_{N2}J_T & \cdots & M_{NN}J_T \end{bmatrix} \right)$$

$$= tr\left(\sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij}J_T M_{ji}J_T \right)$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} tr(M_{ij}J_T M_{ji}J_T) = \sum_{i=1}^{N} \sum_{j=1}^{N} tr(M_{ij}e_T e_T' M_{ji}e_T e_T')$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} (e_T' M_{ij} e_T)^2, \qquad (A4)$$

$tr(M(I_N \otimes J_T)M(J_N \otimes I_T))$

$$= tr\left([I_T \quad I_T \quad \cdots \quad I_T] \begin{bmatrix} M_{11}J_T & M_{12}J_T & \cdots & M_{1N}J_T \\ M_{21}J_T & M_{22}J_T & \cdots & M_{2N}J_T \\ . & . & \cdots & . \\ . & . & \cdots & . \\ M_{N1}J_T & M_{N2}J_T & \cdots & M_{NN}J_T \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} M_{1i} \\ \sum_{i=1}^{N} M_{2i} \\ . \\ . \\ \sum_{i=1}^{N} M_{Ni} \end{bmatrix} \right)$$

$$= tr\left(\begin{bmatrix} \sum_{j=1}^{N} M_{j1}J_T & \sum_{j=1}^{N} M_{j2}J_T & \cdots & \sum_{j=1}^{N} M_{jN}J_T \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{N} M_{1i} \\ \sum_{i=1}^{N} M_{2i} \\ . \\ . \\ \sum_{i=1}^{N} M_{Ni} \end{bmatrix} \right)$$

$$= \sum_{k=1}^{N} tr\left(J_T \left(\sum_{i=1}^{N} M_{ki} \right) \left(\sum_{j=1}^{N} M_{jk} \right) \right)$$

$$= \sum_{k=1}^{N} e_T' \left(\sum_{i=1}^{N} M_{ki} \right) \left(\sum_{j=1}^{N} M_{jk} \right) e_T, \qquad (A5)$$

$tr(M(J_N \otimes I_T)M(J_N \otimes I_T))$

$$= tr\left([I_T \quad I_T \quad \cdots \quad I_T] \begin{bmatrix} \sum_{i=1}^{N} M_{1i} \\ \sum_{i=1}^{N} M_{2i} \\ . \\ . \\ \sum_{i=1}^{N} M_{Ni} \end{bmatrix} [I_T \quad I_T \quad \cdots \quad I_T] \begin{bmatrix} \sum_{i=1}^{N} M_{1i} \\ \sum_{i=1}^{N} M_{2i} \\ . \\ . \\ \sum_{i=1}^{N} M_{Ni} \end{bmatrix} \right)$$

$$= tr\left(\left(\sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij} \right)^2 \right), \qquad (A6)$$

$$\text{vec}(M(I_N \otimes J_T)M')' \text{vec}(M) = tr(M(I_N \otimes J_T)M)$$

$$= tr\left(\left(\sum_{i=1}^{N} M_{ii} \right) J_T \right)$$

$$= e_T' \left(\sum_{i=1}^{N} M_{ii} \right) e_T, \qquad (A7)$$

$$\text{vec}(M(J_N \otimes I_T)M')'\,\text{vec}(M) = tr(M(J_N \otimes I_T))$$
$$= tr\left(\sum_{i=1}^{N}\sum_{j=1}^{N} M_{ij}\right), \qquad (A8)$$

$$\text{vec}(M)'\,\text{vec}(M) = tr(M) = NT - K. \qquad (A9)$$

- MATRIX $Z(M)'R$:

Since $M\hat{v} = MMv = Mv = \hat{v}$, we obtain

$$\text{vec}(M(I_N \otimes J_T)M')'\,\text{vec}(\hat{R}) = tr((I_N \otimes J_T)M\hat{v}\hat{v}'M)$$
$$= tr((I_N \otimes J_T)\hat{v}\hat{v}'). \qquad (A10)$$

This can equivalently be written as

$$tr((I_N \otimes J_T)\hat{v}\hat{v}') = \text{vec}(\hat{N}')'(I_N \otimes J_T)\,\text{vec}(\hat{N}')$$
$$= tr(\hat{N}J_T\hat{N}') = e_T'\hat{N}'\hat{N}e_T$$
$$= \sum_{i=1}^{N}\hat{v}_{i.}^2, \quad \text{where } \hat{v}_{i.} = \sum_{t=1}^{T}\hat{v}_{it}. \qquad (A11)$$

Doing similar computations, we have

$$\text{vec}(M(J_N \otimes I_T)M')'\,\text{vec}(\hat{R}) = tr((J_N \otimes I_T)M\hat{v}\hat{v}'M)$$
$$= tr((J_N \otimes I_T)\hat{v}\hat{v}')$$
$$= \text{vec}(\hat{N}')'(J_N \otimes I_T)\,\text{vec}(\hat{N}')$$
$$= tr(\hat{N}\hat{N}'J_N) = e_N'\hat{N}\hat{N}'e_N$$
$$= \sum_{t=1}^{T}\hat{v}_{.t}^2, \quad \text{where } \hat{v}_{.t} = \sum_{i=1}^{N}\hat{v}_{it}, \qquad (A12)$$

and

$$\text{vec}(M)'\,\text{vec}(\hat{R}) = tr(M\hat{R})$$
$$= \hat{v}'\hat{v} = \sum_{i=1}^{N}\sum_{t=1}^{T}\hat{v}_{it}^2. \qquad (A13)$$